Dean&Francis
ISSN 2959-6157

Cultural Information Extraction and Overseas Dissemination of Online Literature

Chuangwei Bao

College of Information and Management Science, Henan agricultural university, Zhengzhou, Henan,450046, China
baochuangwei214@163.com

Abstract:

Since the millennium, China's online literature has achieved phased results in spreading overseas and cultural dissemination. Chinese online literature is not only a carrier of commercial products such as novels, but also an excellent achievement in spreading Chinese culture. However, despite its excellent performance on foreign online literature platforms, there is neither a comprehensive review of existing literature on cultural dissemination phenomena nor a detailed exercise on how to fully analyze its role. Therefore, this article reviews multiple articles and analyzes and summarizes the theories of cultural information extraction and cultural knowledge graph construction, considers the effectiveness of online literature dissemination worldwide from the perspectives of cultural differences and acceptance, and uses CANDLE, an end-to-end methodology, to extract high-quality cultural information on a large scale. Common Sense Knowledge (CCSK). By utilizing natural language processing (NLP) technology, the emotional resonance of online literature in cross-cultural communication is analyzed, providing theoretical support and practical guidance for enhancing the overseas acceptance of online literature. Finally, a cultural knowledge graph is used to provide basic information for understanding the possible role of Chinese culture in various cultures.

Keywords: Cultural information extraction; Cultural knowledge graph; Cultural globalization; Chinese online literature; NLP.

1. Introduction

In recent years, with the rapid development of the digital age and the deepening of globalization, the cross-cultural dissemination of online literature has gradually attracted widespread attention from the academic community. Online literature is not only an important medium for cultural expression, but also

an important platform for cultural identity and exchange. However, in this context, the cultural distance between different cultures has an increasingly significant impact on the effectiveness of communication. The concept of cultural distance mainly refers to the differences in values, language, customs, and other aspects between cultures, which may lead to misunderstandings, distortions, or even resistance in the process of dissemination. Meanwhile, the application of natural language processing (NLP) technology can help analyze and understand the mechanism by which cultural factors affect communication effectiveness. Therefore, studying the relationship between cultural distance and dissemination effects, and combining NLP to explore the overseas acceptance of online literature, has important practical significance and academic value.

The main purpose of this study is to explore in depth the impact of cultural distance on the overseas dissemination effect of online literature (hereinafter referred to as' online literature '), aiming to determine the dissemination effect of online literature and its impact on other cultures by constructing a national cultural knowledge graph and combining national culture with the dissemination effect of online literature. At the same time, with the help of natural language processing (NLP) technology, the emotional color, language features, and thematic associations of online literature in cross-cultural communication are analyzed, providing theoretical support and practical guidance for enhancing the overseas acceptance of online literature. Specifically, the study will systematically consider the following aspects: firstly, quantifying the relationship between cultural distance and dissemination effects; The second is to explore the differences in audience acceptance of online literature under different cultural backgrounds; The third is to analyze the impact of text features on audience behavior through NLP technology, and ultimately provide suggestions for improving the creation and dissemination strategies of online literature.

The study of the relationship between cultural distance and communication effects not only enriches the theoretical perspectives of communication studies and cultural research, but also provides a necessary theoretical basis for promoting the international development of online literature. At the cultural level, in-depth exploration of how cultural distance affects the core dissemination effect of online literature can help understand the audience's interpretation and acceptance mechanisms of texts in different cultural backgrounds, which has important theoretical value for cross-cultural communication. In addition, utilizing natural language processing techniques to analyze text characteristics will provide new perspectives and methodological support for the development of digital humanities, further promoting interdisciplinary research in computational linguistics and cultural communication. At the same time, the results of this study will also provide effective strategic recommendations for online literature creators, helping them better adapt to the international market and enhance the international competitiveness and influence of their works.

The theoretical framework of this study is mainly based on cross-cultural communication theory, cultural distance theory, and data-driven natural language processing technology. Firstly, cross-cultural communication theory emphasizes the challenges and opportunities faced in communication under different cultural backgrounds, while cultural distance theory provides an important perspective for understanding how cultural differences affect communication effectiveness. On this basis, combined with natural language processing technology, cultural elements and emotional characteristics in online text can be analyzed in more detail, revealing how cultural factors are transmitted through text and affect the audience's psychological perception and behavioral responses. Specifically, the study will use quantitative methods to evaluate multiple dimensions of cultural distance, including language habits, values, and social norms, while applying NLP technology for deep mining of sentiment analysis, topic modeling, and keyword extraction in online literature. This multidimensional theoretical framework aims to establish an empirical relationship between cultural distance and communication effects, providing solid theoretical support for subsequent research.

2. Cultural Information Extraction

Structured knowledge is crucial for many AI applications. For human centered artificial intelligence, essential common sense knowledge is currently only covered in a few structured knowledge projects. This article uses CANDLE, an end-to-end methodology aimed at extracting high-quality cultural common sense knowledge (CCSK) on a large scale [1]. CANDLE extracts CCSK from a vast network corpus and organizes it into coherent groups, including three domains (geography, religion, profession) and multiple cultural aspects (food, drink, clothing, tradition, ceremony, behavior). To understand CCSK, it is necessary to consider different cultural backgrounds. For example, employees working in European factories must have different cultural knowledge and understanding from those working in the Middle East or East Asia. Through AI and NLP, different CCSKs can be constructed based on different cultural backgrounds to evaluate the varying effects of online cultural dissemination abroad.

ISSN 2959-6157

2.1 Methods for Extracting Cultural Information

Almost all LM are buried in massive amounts of information, appearing in latent form and in billions of parameter forms. At the same time, knowledge cannot be hidden, so it can only be observed through LM based outputs in applications. Sometimes cultural information is very important, and even large LMs like GPT-3 may perform poorly. For example, if one does not understand that dragons are symbols of luck and majesty in East Asia, it is very likely that dragons will be compared and used together with real animals in future applications, such as keeping dragons in cages for people to watch. However, common sense knowledge tells people that dragons do not exist in reality, so all knowledge inferred based on the common sense of life that dragons exist is wrong.

So it is necessary to design an end-to-end methodology and system, called CAN-DLE (Large scale Extraction of Cultural Knowledge), in order to extract and systematically organize CCSKs from massive amounts of information. The output of CANDLE is a set of 1.1 million CCSKs, organized into 60000 coherent clusters. This set of data is divided into three domains: geography, religion, and occupation, with a total of 386 instances referred to as subjects (or cultural groups). Each subject corresponds to five aspects of culture: food, drink, clothing, ceremony, tradition (for geography and religion), or behavior (for profession). CANDLE requires six steps to operate. Firstly, use simple techniques for subject detection (named entity recognition NER and string matching). Secondly, identify candidate claims based on rule-based filtering. Thirdly, categorize the claims into specific cultural levels. Fourth, use the most advanced technology for claim clustering, and fifth, adopt a simple and effective method to extract concepts from claims. Finally, evaluate the interestingness of the claim by combining multiple features such as frequency, specificity, uniqueness, etc. In this way, people can avoid generating overly vague claims and focus more on precise and accurate ones [2].

2.2 Result of Cultural Information Extraction

Collect CCSK to cover three cultural domains: geography, religion, and occupation. In the field of geography, it is subdivided into four sub domains: countries, continents, geopolitical regions (such as the Middle East, Southeast Asia, etc.), and states in the United States. These data come from the GeoNames database, which also provides aliases. These aliases can be further enriched by geographic terms from Wikipedia.

Various aspects of culture. Mainly consider five aspects: food, beverages, clothing, rituals and traditions (applicable to geography/religion) or behavior (applicable to profession). Execution status and result statistics can be used to adjust system parameters for small hidden data.

3. Introduction to Cultural Studies

3.1 Construction of Cultural Knowledge Graph

The ethnic cultural knowledge graph is established through knowledge extraction and manual construction, which includes 1064 knowledge triplets [3]. Firstly, obtain ethnic cultural text data such as Baidu Baike and Wikipedia through web crawling technology; Then, named entity recognition technology and relationship extraction technology are used to obtain entity and relationship information in ethnic texts, and triplets are constructed based on (entity, relationship, entity) or (entity, attribute, entity) methods and manually proofread. Finally, they are stored in a database. The construction of the ethnic cultural knowledge graph mainly includes three aspects: the construction of the category system of the ethnic cultural knowledge graph, the definition of ethnic cultural entity attributes, and the definition of ethnic cultural relationship attributes.

3.2 Construction of Category System for Ethnic Cultural Knowledge Graph

The construction of the category system for the ethnic cultural knowledge graph is the basic work of cultural knowledge graph construction. Through detailed division of cultural data, ethnic cultural text data is divided into six categories, as shown in Table 1.

According to Table 1, each ethnic group is divided into a system represented by six general categories, which not only supports the establishment of a cultural knowledge graph, but also connects to CANDLE's end-to-end methodology.

Table 1. Classification of Ethnic Cultural Categories.

Category Name	Category Name
Religious culture	Primitive religious culture, Buddhist culture
dietary culture	Beverage culture, snack culture
architectural culture	Residential architectural culture, temple architectural culture

dress culture	Women's clothing culture, men's clothing culture
Customs and Culture	Festival culture, funeral culture, marriage culture
art and culture	Music culture, dance culture, instrument culture, handicraft art culture

After determining the classification system of ethnic culture, it is necessary to define the attributes related to cross-border ethnic culture based on each category, including the entity's name, alias, description content, entity label, and some characteristics of the entity's existence. By defining this information of the entity, the entity can provide a complete and detailed description of cross-border ethnic culture. For the entity of "Dragon Statue", its entity label category information is "Chinese Nation", "Chinese Ethnic Customs and Culture", "Chinese Ethnic Festival Culture", etc. Establishing the relationship between entities can integrate and associate knowledge in the field of ethnic culture knowledge graph, making the knowledge graph of ethnic culture more representative and enhancing the query performance of the knowledge graph of ethnic culture [4]. The entity relationships in the field of ethnic culture are complex and intricate. It can be mainly summarized as: inclusion, location, genus, attribute. Finally, by combining encyclopedia entry information with structured knowledge, the information of the knowledge triplet can be obtained.

3.3 Ethnic Text Classification Model

In the process of collecting data, it is important to realize that not all data can express the desired results well. For example, some data may provide vague information about how people in a certain region celebrate a certain festival every year, which is a form of cultural information. However, the system is not clear about which festival it is, so it is necessary to filter cultural information during the collection process.

There are also a large number of professional terms in ethnic cultural data, and ordinary word segmentation tools do not have an ethnic cultural lexicon, so the segmentation effect is poor, which in turn affects subsequent classification work. For example, 'Dai people believe in Theravada Buddhism', and the result obtained without using the cross-border ethnic cultural lexicon is' Dai people believe in Theravada Buddhism '. However,' Theravada Buddhism 'is a professional term that cannot be separated, and there are many similar words in cross-border ethnic cultures. To this end, Chen et al. used a manually constructed cross-border ethnic culture lexicon to assist jieba tool in word segmentation, and trained the existing cross-border ethnic culture data with the Word2Vec model to obtain word vectors,And use cross-border ethnic word vectors to obtain the $w_i, i \in [1, N]$ embedding representation $x_i \in R^{100}$ of each word in the title, where N represents the number of words in the title [5]. They use bidirectional gated recurrent units (Bi GRU) to obtain word level representations of titles. Bi GRU contains hidden vector representations for titles x_1 to x_N obtained from the forward GRU and hidden vector representations for titles x_N to x_1 obtained from the backward GRU.

$$h_i = BiGRU(x_i), i \in [1, N] \#(1)$$

Among them, hi represents the combination of the forward and backward hidden state information of the i word in the title, $h_i = [\underset{h_i}{\rightarrow}, \underset{h_i}{\leftarrow}]$. The experimental results show that this method can significantly improve the effect of word segmentation, making the words under different cultures and semantics easier to understand and classify, and greatly enhancing the efficiency of word segmentation.

4. Current limitations and Future Prospects

Cultural distance and dissemination effects: A study on the overseas acceptance of online literature based on NLP [6]. In the context of globalization, the cross-border dissemination of cultural products has become an important research field. The theme of "Research on the Overseas Acceptance of Online Literature Based on NLP: Cultural Distance and Communication Effect" aims to explore the impact of cultural distance on the overseas dissemination effect of online literature, and to use natural language processing technology for data analysis and case studies. By analyzing relevant literature, this article will systematically elaborate on the current situation, problems, and potential research value in this field.

4.1 Existing limitations

Experience and Problems of Overseas Dissemination of Fantasy Online Articles: A Study Based on Overseas Reader Reviews of Amazon's "Panlong" [7].

This paper explores the acceptance of the fantasy web article "Panlong" among overseas readers, using qualitative analysis and reader comments to explore its cultural dissemination experiences and challenges. Research has

ISSN 2959-6157

shown that cultural distance is an important factor affecting readers' understanding and acceptance. The data shows that readers' acceptance of cultural elements is closely related to their understanding and interest in Chinese culture. This paper provides empirical case support for this study, indicating that the greater the cultural distance, the lower the acceptance.

However, simply understanding Chinese culture is not enough. It is also necessary to recognize the differences in the cultures of various ethnic groups. For example, if a certain ethnic group believes that dragons represent evil symbols and dislikes the appearance of dragon elements in online literature, even if the quality of online literature is good, it will not be beneficial for spreading Chinese culture.

Therefore, based on the knowledge graph of ethnic culture and the research on the overseas dissemination of online literature under NLP, there are limitations.

4.2 Future Prospects

A study on the dissemination effect of Chinese media on overseas social media platforms from a distance perspective [8]. This article explores the dissemination effect of Chinese media on overseas social media from the perspective of communication studies, emphasizing the role of cultural distance in information dissemination. Through quantitative analysis and data mining, the paper reveals how cultural differences create barriers to understanding in information dissemination. This study provides an important theoretical basis for the selection of communication channels and cultural adaptation strategies for this topic.

A Study on the Overseas Acceptance of the Film "Big Fish and Begonia" [9], this article focuses on the overseas market performance of the film "Big Fish and Begonia", and explores its acceptance on the international stage by analyzing the relationship between audience feedback and cultural belonging [10]. The results show that the success of the film lies in its cultural intermediary role and universal value in storytelling. This study provides a reference for the overseas promotion of online literature: how to reduce cultural distance and improve acceptance through narrative strategies in cultural dissemination is a topic worthy of in-depth research.

From the perspective of translation and communication studies, this article examines the cultural phenomenon of "online literature going global" and analyzes the cultural adaptability issues in the translation of Chinese online literature. The paper explores the complex relationship between translation strategies and cultural distance, emphasizing the importance of adopting flexible translation strategies in cross-cultural communication to enhance content acceptance. This perspective has important theoretical reference significance for studying the overseas dissemination strategies of online literature, especially how to use NLP technology for text analysis and translation optimization.

A study on the overseas dissemination effect of Chinese online novels based on audience analysis [11]. This article systematically explores the dissemination effect of online novels overseas through audience analysis methods. Research shows that cultural background and audience psychological expectations greatly affect the reading experience and acceptance of works. Based on data analysis, the author proposes customized communication plans for audiences with different cultural backgrounds. This contribution provides practical guidance for this study on how to optimize the promotion strategy of online literature based on cultural distance and user preferences.

5. Conclusions

In summary, relevant literature has laid a solid theoretical foundation for this study. In the discussion of translation and communication studies, this paper see the crucial role of cultural adaptability in text acceptance, and through audience analysis, people can gain a deeper understanding of the impact of cultural differences on communication strategies. Subsequent research will focus on how to use NLP technology to quantitatively analyze the impact of these cultural distances, in order to provide more scientific strategies for the international dissemination of cultural products.

In the following research, this paper will further explore the application of NLP technology in cross-cultural communication, revealing the specific impact of cultural differences on the promotion of online literature through methods such as big data analysis of emotion recognition and the extraction of cultural information and the construction of a category system for ethnic cultural knowledge graphs. At the same time, this paper will also pay attention to the acceptance and dissemination effects of literary works in different cultural backgrounds, and explore how to enhance the exposure and influence of works through personalized recommendation algorithms and social media marketing strategies. In addition, this paper will also focus on the changing trends of user preferences, analyze the preferences and needs of different groups for online literary works, in order to develop more accurate and effective promotion strategies. Through continuous practice and improvement, this paper believe that more targeted and actionable guidance can be provided for the international dissemination of cultural products, promot-

ing the dissemination and development of online literature on a global scale. In future research, this paper will continue to explore cutting-edge issues in this field, aiming to reveal the deep impact of cultural distance and user preferences on the dissemination of online literature, and provide new ideas and methods for academic research and practical applications in related fields.

References

[1] Zheng Huanzhao, Kong Lingqin, Yan Chen. Experience and problems of overseas dissemination of fantasy online literature: A study based on overseas reader reviews of Amazon's «Panlong». Network Literature Research, 2023, (01): 197-211.

[2] Li De. Value expectations and emotional connections: A study on the impact mechanism of Chinese culture overseas dissemination: An empirical analysis based on overseas audience acceptance intentions. Science and Technology Communication, 2023, 15(18): 65-68.

[3] Chen Gong. Crossing the emotional and cultural divide: A study on the acceptance of international communication audiences. Modern Communication (Journal of Communication University of China), 2021, 43(02): 72-77.

[4] Liao Lisheng. Overseas communication and education of Chinese culture: A review of «Research on Overseas Communication of Chinese Culture». Chinese Journal of Education, 2022, (11): 115.

[5] Fang Wei, Gu Renjie. Research on Wang Xiaobo's acceptance of overseas communication. Literary Competition, 2022, (11): 107-115.

[6] Fan Yugang. Overseas dissemination of Chinese culture and expansion of overseas cultural interests. Journal of Social Sciences, 2023, (03): 191-203.

[7] Wu Yanxia, Jia Hui. Research on the overseas acceptance of the film 'Big Fish & Begonia'[A]//Innovative Research on Foreign Language Education and Translation Development, 2022: 381-388.

[8] Wang Wanbo. The cultural internationalization and overseas dissemination practice of Chinese online literature. Literary Review, 2020, (04).

[9] Chen Chunji, Mao Cunli, Lei Xiongli. Cross border ethnic text classification method integrating domain knowledge graph. Journal of Shanxi University (Natural Science Edition), 2022, 45(04).

[10] Mao Cunli, Wang Bin, Lei Xiongli. Cross border ethnic cultural classification based on domain knowledge graph. Small and Micro Computer System, 2022, 43(05).

[11] Nguyen T P, Razniewski S, Varde A. Extracting cultural commonsense knowledge at scale, in Proceedings of the ACM Web Conference 2023, 2023: 1907-1917.